EGFR Dinucleotide Repeat Polymorphism as a Prognostic Indicator in Non-small Cell Lung Cancer

Sarita Dubey, MD,[*] Patricia Stephenson, ScD,[†‡] Donna E. Levy, MS,[†] Judith A. Miller, BS,[§] Steven M. Keller, MD,[‖] Joan H. Schiller, MD,[¶] David H. Johnson, MD,[#] and Jill M. Kolesar, PharmD,[§] for the Eastern Cooperative Oncology Group

Background: The epidermal growth factor receptor (EGFR) has been implicated in tumor growth and progression. Intron 1 of the *EGFR* gene contains a polymorphic simple sequence repeat (SSR) of 14 to 21 CA dinucleotides, the length of which correlates inversely with the level of EGFR transcription. The authors hypothesized that a shorter length of tumor SSR would be associated with poorer survival in patients with non-small cell lung cancer (NSCLC).

Methods: Patients enrolled in Eastern Cooperative Oncology Group E3590 (a randomized, prospective trial of adjuvant therapy following resection of stages II and IIIa NSCLC) were randomized to radiation or radiation plus chemotherapy. Genomic DNA extracted from resected tumors was amplified for EGFR intron 1 by polymerase chain reaction and sequenced in a 3730XL DNA analyzer.

Results: One hundred fifty-seven primary tumors were sequenced, 106 (68%) of which were heterozygous for intron 1. The most common genotypes were allele lengths of 17/19 dinucleotides (17.8%), 17/18 (11.4%), and 19/19 (11.4%). Allele status (homozygous versus heterozygous) did not correlate with race, gender, weight, performance status, histology, stage, or survival. Shorter allele length ($\leq$18 versus $>$18 CA dinucleotide repeats) was associated with squamous cell histology ($p = 0.03$). Allele sum of greater than 35 was associated with improved overall survival (log-rank $p = 0.03$, hazard ratio = 0.66).

Conclusion: This is the first study to characterize the EGFR intron 1 SSR polymorphism in NSCLC. Tumors were most commonly heterozygous for SSR length. Squamous histology was associated with a shorter SSR. Longer sequences are associated with improved survival.

Key Words: Intron 1, Simple sequence repeat, Allele

(*J Thorac Oncol.* 2006;1: 406–412)

*University of California, San Francisco, San Francisco, California; †Dana Farber Cancer Institute, Boston, Massachusetts; ‡Harvard School of Public Health, Boston, Massachusetts; §University of Wisconsin-Madison, School of Pharmacy, Madison, Wisconsin; ‖Albert Einstein College of Medicine, Bronx, New York; ¶University of Wisconsin Comprehensive Cancer Center, Madison, Wisconsin; and #Vanderbilt University, Nashville, Tennessee.

The work reported in this article was supported by a National Cancer Institute K12 grant.

Address for correspondence: Sarita Dubey, M.D., University of California, San Francisco, 1600 Divisadero Street, Box 1770, San Francisco, CA 94115; email: sdubey@medicine.ucsf.edu.

ISSN: 1556-0864/06/0105-0406

Lung cancer remains the leading cause of cancer-related deaths in the United States, with 160,000 new cases estimated in the year 2006.[1] Clinical parameters such as stage, performance status, and weight loss reliably predict survival. Although the mutational status of oncogene *TP53* has not correlated with prognosis, *KRAS* mutation has been associated with poor outcome in patients treated with epidermal growth factor receptor (EGFR) inhibitors.[2,3]

The signaling pathways triggered by EGFR activation render the EGFR a logical correlate with tumor growth and patient prognosis. Evidence for an important role of EGFR in neoplasia includes its presence in premalignant bronchial epithelium[4,5] and decreased expression following regression of metaplasia. EGFR overexpression in resected non-small cell lung carcinoma (NSCLC) specimens has been associated by some investigators with a worse prognosis.[6] Others, however, found no statistically significant relationship, but did note a trend toward worse prognosis with elevated EGFR or increased *EGFR* gene copy.[7] *EGFR* mutations have been identified as a prognostic indicator with improved survival in patients treated with chemotherapeutic agents.[3]

Transcription of the *EGFR* gene is regulated by two enhancer elements. The first is located upstream near the transcription initiation site and the second downstream in intron 1.[8] The latter contains a characteristic simple sequence repeat (SSR) of CA dinucleotides, the length of which exhibits interethnic differences.[9] In addition, the length of this polymorphism has been shown to correlate inversely with EGFR transcription in both in vitro experiments and breast cancer specimens.[10,11] Eastern Cooperative Oncology Group (ECOG) 3590 (a randomized, prospective trial of adjuvant therapy following resection of stages II and IIIa NSCLC) provided a unique opportunity to characterize the frequency of *EGFR* polymorphisms in NSCLC specimens and determine whether the length of the SSR correlated with survival.

PATIENTS AND METHODS

Patient Population

ECOG 3590 was a multicenter, randomized, prospective trial designed to compare concurrent chemotherapy plus radiation to radiation alone in the adjuvant treatment of NSCLC.[12] Patients with completely resected stages II and IIIA were randomized to receive either thoracic radiation (50.4 Gy) or thoracic radiation concurrent with chemother-

apy. Chemotherapy consisted of four 28-day cycles of cisplatin, 60 mg/m^2 on day 1, and etoposide, 120 mg/m^2 on days 1 to 3. Survival was similar in both study arms. Patients who participated in ECOG 3590 and whose tumor tissue was available for further analysis were eligible for enrollment in a laboratory-based investigation (E4592).[2] Institutional review board approval of the laboratory study was obtained at each institution. Informed consent was obtained from the patient or surrogate.

Polymerase Chain Reaction

Genomic DNA extracted from resected tumors was stored at −80°C. Intron 1 of the *EGFR* gene was amplified by polymerase chain reaction (PCR) using primers purchased from Integrated DNA Technologies, Inc. (Coralville, IA). One microliter of DNA was mixed with 1 μl of forward primer, 5′-GGA TGT CTC TTT TTG CTG TT, 1 μl of reverse primer 5′-CTG TCA AGT GGG TTT ATG GT, 7 μl of nuclease-free water, and 10 μl of Promega PCR Master Mix (Promega Corporation, Madison, WI). Promega PCR Master Mix contained *Taq* DNA polymerase in reaction buffer with pH 8.5, nucleotides (dATP, dGTP, dCTP, dTTP), and $MgCl_2$. An MJ Research (PTC 100) thermocycler (Bio-Rad Laboratories, Inc., Waltham, MA) was programmed to cycle 55 times through the denaturation at 95°C, annealing at 50°C, and extension at 72°C, producing a 185 to 197 base pair product, depending on the number of repeats.

The PCR product was purified with ExoSAP-IT (exonuclease I and shrimp alkaline phosphatase in a specially formulated buffer) from USB Corporation (Cleveland, OH), as recommended by the manufacturer. Five microliters of PCR product was mixed with 2 μl of ExoSAP-IT and incubated at 37°C for 15 minutes in a thermal cycler. ExoSAP-IT was inactivated by heating to 80°C for 15 minutes.

Sequencing

The cleaned product was combined with the forward PCR primer and sequenced at the University of Wisconsin Biotechnology Center (Madison, WI). DNA PCR product consisting of 20 ng and forward primer (5 pmol) were added to the Enzyme Reaction Mix (2 μl BigDye plus 3 μl buffer and 1 μl of dimethyl sulfoxide) and water to 20 μl. Cycling conditions were as follows: 96°C for 3 minutes, then 35 cycles of 96°C for 10 seconds, 50°C for 15 seconds, 60°C for 3 minutes, finished with 72°C for 7 minutes.

Cleanup of sequenced product was performed with CleanSeq beads as recommended by the manufacturer (Agencourt Bioscience Corporation, Beverly, MA). Samples were electrophoresed on an Applied Biosystems (Foster City, CA) 3730xl automated DNA sequencing instrument. Data were analyzed using PE-Biosystems version 3.7 of Sequencing Analysis. The resulting chromatograms were directly sequenced and scored by using computer software Chromas version 2.3 (Griffith University, Queensland, Australia) (Figure 1). In our experience, the quality of the sequence is more dependent on the location of the sequencing primer than on the orientation (forward or reverse) of the primer. In this assay, the primer was sufficiently downstream of the repeats to provide accurate reads in the forward orientation. To ensure accuracy, sequencing results were evaluated by two independent observers and inconclusive results were repeated. Inconclusive results were caused by an inadequate PCR amplification that resulted in a failed sequencing run.

Statistical Analysis

Tumor samples were analyzed for both allele status and length. Intron 1 was classified as homozygous when the two alleles had the same CA dinucleotide repeat number. In a population-based case control study, the presence of two long alleles of 19 or more was associated with an increased risk of breast cancer.[13] In a preclinical head and neck cancer study, a difference in response to the EGFR tyrosine kinase inhibitor erlotinib was seen when the sum of CA dinucleotides in both alleles was dichotomized into >35 or ≤ 35.[14] Because a clinical difference was seen in different studies using different cutoff values, both cutoff values were explored for our analysis. The mean SSR length was calculated for each patient and dichotomized as short (≤ 18) or long (>18) based on the number of dinucleotide repeats. Samples were also divided into two cohorts based on the sum of their alleles: less than or equal to 35 versus greater than 35. Recursive partitioning and regression-tree techniques were used on the data for each allele, with overall survival as the endpoint of interest. This method attempts to mathematically divide the data into appropriate homogeneous partitions (i.e., exploring for a cutpoint or cutpoints). For alleles 1 and 2, no partitions were found when the covariate was treated as a continuous

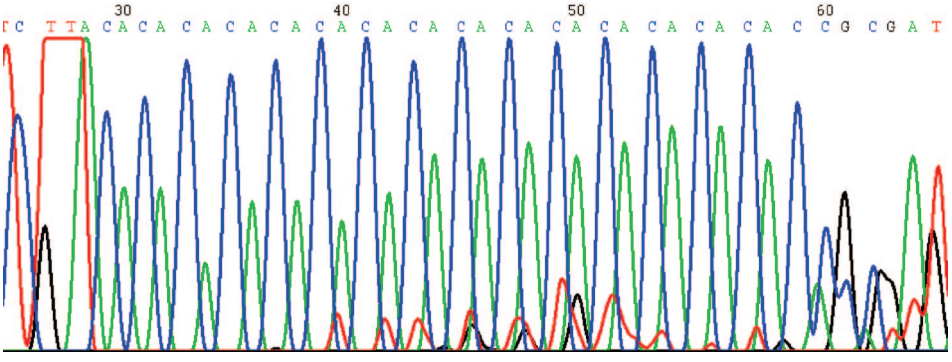

FIGURE 1. Intron 1 CA repeat direct sequencing by Chromas version 2.3.

TABLE 1. Correlation of Baseline Characteristics with Allele Status and Length

Baseline Characteristics	E3590 Study (%)	EGFR Dinucleotide Repeat Polymorphism Study (%)						
		Total	Homozygous Alleles	Heterozygous Alleles	*p* Value	≤18 $(CA)_n$	>18 $(CA)_n$	*p*Value
Total	488	157						
Age								
> 60	243 (49.7)	85 (54.1)	28 (54.9)	57 (53.8)	>0.9	47 (56.6)	38 (51.3)	0.52
≤ 60	245 (51.3)	72 (45.9)	23 (45.1)	49 (46.2)		36 (43.4)	36 (48.7)	
Gender								
Male	285 (58.4)	95 (60.5)	26 (51.0)	69 (65.1)	0.12	51 (61.5)	44 (59.5)	0.87
Female	203 (41.6)	62 (39.5)	25 (49.0)	37 (34.9)		32 (38.5)	30 (40.5)	
Race								
Caucasian	418 (85.6)	138 (87.9)	46 (90.2)	92 (86.8)	0.61	77 (92.8)	61 (82.4)	0.05
Other	70 (14.4)	19 (12.1)	5 (9.8)	14 (13.2)		6 (7.2)	13 (17.6)	
ECOG PS								
0	190 (38.9)	60 (38.2)	19 (37.2)	41 (38.7)	>0.9	31 (37.3)	29 (39.2)	0.87
1	297 (60.8)	97 (60.8)	32 (62.8)	65 (61.3)		52 (62.7)	45 (60.8)	
Weight loss in 6 mo								
<5%	384 (78.6)	125 (79.6)	38 (74.5)	87 (82.1)	0.29	64 (77.1)	61 (82.4)	0.43
≥5%	104 (21.3)	32 (20.4)	13 (25.5)	19 (17.9)		19 (22.9)	13 (17.6)	
Stage								
II	202 (41.3)	75 (47.8)	25 (49.0)	50 (47.2)	0.87	39 (47)	36 (48.7)	0.87
IIIa	285 (58.4)	82 (52.2)	26 (51)	56 (52.8)		44 (53)	38 (51.3)	
Histology								
Adenocarcinoma	259 (53.1)	83 (52.9)	27 (52.9)	56 (52.8)	>0.9	40 (48.2)	43 (58.1)	0.04
Squamous	170 (34.8)	53 (33.8)	17 (33.3)	36 (34.0)		35 (42.2)	18 (24.3)	
Other	59 (12.0)	21 (13.3)	7 (13.7)	14 (13.2)		8 (9.6)	13 (17.6)	

ECOG PS, Eastern Cooperative Oncology Group Performance Status; 0, normal activity; 1, symptoms of disease but ambulatory; $(CA)_n$; CA dinucleotide repeat.

variable modeling survival using methods of Breiman et al. in the statistical software R.[15] When the sum of the alleles was recursively partitioned, two homogeneous groups were found (≥35.5 versus ≤35.5).

Descriptive statistics were used to describe patient baseline characteristics. Using Fisher's exact test at the 0.05 level, the association of both allele status and allele length with patient characteristics was investigated.

Overall survival was defined as the time from randomization until death or date last known alive. Survival curves were estimated by the method of Kaplan and Meier and compared using log-rank tests. Cox proportional hazards models were used to investigate the relationship between survival and allele status after adjusting for various prognostic factors. Similar models were also constructed for allele length. Stepwise selection was used to choose parsimonious models considering the following factors: gender (male versus female), age (>60 versus ≤60 years), histology (squamous versus other), weight loss in the previous 6 months (<5% versus ≥5%), stage (II versus IIIa), performance status (1 versus 0), and race (Caucasian versus other). All reported

TABLE 2. Frequency of CA Dinucleotide Repeats in Homozygous Samples

Homozygous (n = 51)	
$(CA)_n$	No. (%)
15	1 (0.6)
16	2 (1)
17	3 (2)
18	10 (6.3)
19	18 (11.4)
20	12 (8)
21	4 (2.5)
22	1 (0.6)

Percentages are calculated based on the total number (N=157)

TABLE 3. Frequency of CA Dinucleotide Repeats in Heterozygous Samples*

Heterozygous n = 106			
$(CA)_n$	No. (%)	$(CA)_n$	No. (%)
11/13	1 (0.6)	17/20	4 (2.5)
13/15	1 (0.6)	18/19	4 (2.5)
15/16	1 (0.6)	18/20	4 (2.5)
15/17	3 (2)	19/20	9 (5.7)
16/17	9 (5.7)	19/21	1 (0.6)
16/18	4 (2.5)	20/21	2 (1)
16/19	2 (1)	20/22	9 (5.7)
17/18	18 (11.4)	21/22	3 (2)
17/19	28 (17.8)	21/23	3 (2)

*Percentages are calculated based on the total number (n = 157).

p values are associated with two-sided tests. Logistic regression models were used to assess the association between the mean length of SSR with disease histology.

To assess for possible evolutionary changes occurring in this allele polymorphism, Hardy-Weinberg analysis was performed by matching allele frequencies in the observed individuals to those expected under Hardy-Weinberg equilibrium (HWE). Genotypes were dichotomized as short or long using three different cutoffs (<17 versus ≥ 17), (<18 versus ≥ 18), and (<19 versus ≥ 19), and compared using a χ^2 goodness-of-fit test.

RESULTS

Four hundred eighty-eight patients were enrolled in the E3590 trial. One hundred fifty-seven primary tumor samples were available for *EGFR* intron 1 analysis. The demographic distribution of these patients was similar to the patients on the ECOG 3590 trial (Table 1). The lengths of the two alleles were heterozygous in 106 samples (68%). The most common genotypes contained allele lengths of 17/19 dinucleotides (17.8%), 17/18 (11.4%), and 19/19 (11.4%) (Tables 2 and 3).

Hardy-Weinberg Equilibrium Analysis

Allele frequencies were dichotomized as short or long using three different cutoffs (<17 versus ≥ 17, <18 versus ≥ 18, and <19 versus ≥ 19). For all cutoffs used, the alleles were not in HWE (<17 versus ≥ 17, $p < 0.0001$; <18 versus ≥ 18, $p < 0.034$; and <19 versus ≥ 19, $p < 0.034$). The observed frequency of the heterozygous pair (short/long) was less than expected, whereas the observed frequencies of the homozygous pairs (short/short) and (long/long) were greater than expected. When analyzed by race, alleles were not in HWE in Caucasians for each of the cutoffs ($p < 0.001$, $p = 0.03$, and $p < 0.001$, respectively). In African Americans, alleles were in HWE for cutoffs less than 18 versus greater than or equal to 18 ($p = 0.46$) and less than 19 versus greater than or equal to 19 ($p = 0.46$) However, once again, the frequencies of the heterozygous pairs were fewer than expected and the frequencies of homozygous pairs exceeded the expected frequency.

Analysis of Allele Status

Allele status (homozygous versus heterozygous) was not significantly associated with baseline characteristics such as race, gender, weight loss in previous 6 months, performance status, histology (squamous versus other, $p > 0.9$; adenocarcinoma versus other, $p > 0.9$), and stage (Table 1). In addition, there was no significant difference in median survival between homozygous and heterozygous patients (50.7 months versus 38.5 months, $p = 0.40$) (Table 4 and Figure 2). In multivariate analyses, allele status was not associated with survival.

Analysis of SSR Length

Seventy-four of the 157 tumor samples (47%) were categorized as having long (>18) allele length (Table 1). A logistic regression model with length as the outcome and histology (squamous, other, and adenocarcinoma) was conducted. There was a significant association between SSR length and histology ($p = 0.047$). Patients with squamous histology had approximately two times the odds of a shorter SSR length in comparison with patients with adenocarcinoma histology (odds ratio, 2.1; 95% confidence interval [CI], 1.02–4.27; $p = 0.01$). Patients with other histology had lower odds of shorter SSR length in comparison with patients with squamous cell histology, although these two groups were not statistically different from each other (odds ratio, 0.66; 95% CI, 0.25–1.76; $p = 0.11$). Caucasian race was associated with a shorter SSR length than other races ($p = 0.05$). SSR average length was not significantly associated with gender, weight loss, stage, or performance status.

The median survival of the 488 patients treated on clinical trial E3590 was 39 months (95% CI, 30–52 months). The median survivals of those who were randomized to

TABLE 4. Correlation of Median Survival with Allele Status and Length

	No. of Patients (%)	Median survival (mo)
Allele status		
Homozygous	51 (32)	50.7 ($p = 0.40$)
Heterozygous	106 (68)	38.5
Sum of alleles $(CA)_n$		
≤35	45 (28.6)	29.2 ($p = 0.03$)
>35	112 (71.3)	41

$(CA)_n$, CA dinucleotide repeats.

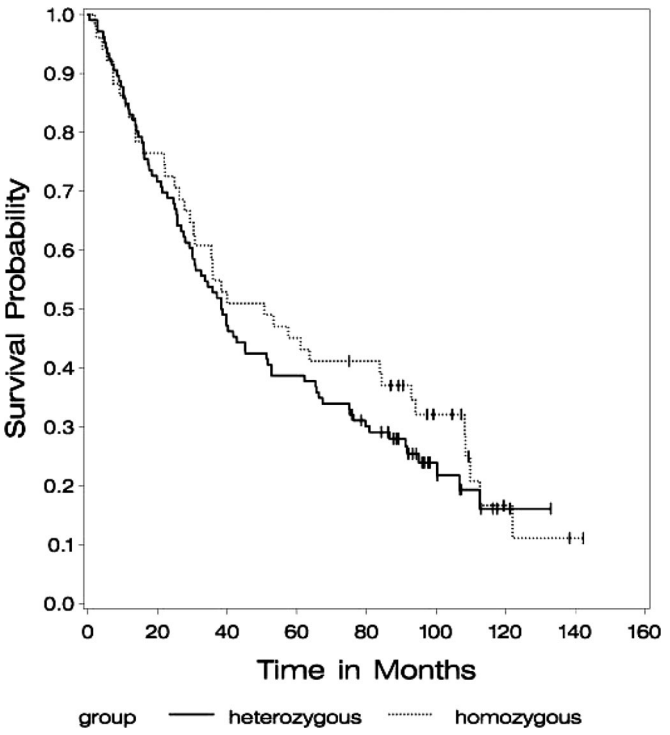

FIGURE 2. Survival did not correlate with allele status. There was no statistically significant difference between the median survival of homozygous (50.7 months) and heterozygous (38.5 months) alleles ($p = 0.40$).

postoperative radiation and postoperative chemoradiation were 39 and 38 months, respectively ($p = 0.56$). Of the 157 patients who were analyzed for intron 1 polymorphism, using multivariate analysis with Cox proportional hazards, when analyzed by the sum of the alleles, a significant difference was found in favor of the longer allele. Patients with an allele sum of less than or equal to 35 had a median overall survival of 29.2 months (95% CI, 21.3–51.4 months), whereas patients with an allele sum of greater than 35 had a median overall survival of 41.0 months (95% CI, 35.8–66.3) (hazard ratio [HR], 0.66; $p = 0.03$) (Figure 3 and Table 4). When controlling for treatment randomization in this study, the difference in overall survival is still statistically significant between the two allele groups (χ^2 $p = 0.03$; HR, 0.655), with the longer allele being associated with higher survival (Table 5). Exploratory analyses using recursive partitioning was conducted on the allele length averages. Two homogeneous groups were found (<18 versus ≥18) whose survivals were identical to homogeneous groups of the sum of the allele lengths (≤35 versus <35).

DISCUSSION

The cell surface receptor family is representative of the receptor tyrosine kinases and consists of four subfamilies. The most extensively studied of the subfamilies have been

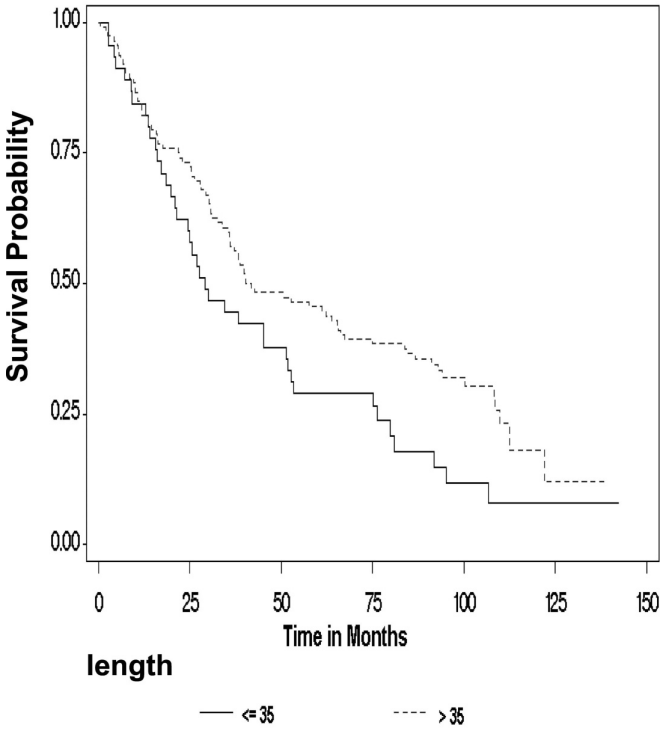

FIGURE 3. Survival correlated with the number of dinucleotide repeats when dichotomized by mean sum of alleles. Allele sum of less than or equal to 35 had a median overall survival of 29.2 months (95% CI, 21.3–51.4 months), whereas patients with an allele sum greater than 35 had a median overall survival of 41.0 months (95% CI, 35.8–66.3 months; HR, 0.66; $p = 0.03$).

TABLE 5. Survival by Treatment Groups

No.	$(CA)_n$ Sum of Both Alleles	Treatment	Median OS (mo)	95% CI
23	≤35	RT	38.2	21.0–52.7
56	>35	RT	41.5	32.5–92.9
22	≤35	Chemotherapy and RT	25.3	16.0–75.2
56	>35	Chemotherapy and RT	40.9	30.8–75.0

$(CA)_n$, CA dinucleotide repeats; RT, radiation therapy.

EGFR (ERBB1) and Her2-neu (ERBB2). Both molecules are expressed in normal and neoplastic tissue. On ligand activation of the extracellular domain, homodimerization between similar subfamilies or heterodimerization between different subfamilies occurs. This results in phosphorylation of the intracellular tyrosine kinase domain, which in turn triggers downstream pathways including mitogen-activated protein kinase, phosphatidyl inositol-3 kinase, and Akt.[16] These processes, which then regulate several transcription factors involved in cell growth, migration, adhesion, and apoptosis, contribute to the mitogenic potentials of cells and thus tumor growth. Because EGFR-dependent growth increases following neoplastic transformation,[17] the level of EGFR expression would be expected to correlate with survival.

Indeed, breast tumors expressing Her2-neu were associated with shorter survival.[18] In addition, the level of Her2-neu expression correlated with response to anti–Her2-neu treatment.[19] However, in patients with NSCLC, tumor EGFR protein expression does not reproducibly correlate with survival or treatment response to EGFR inhibitors.[7,20] Theoretical but as yet unproven explanations for the lack of correlation include the ability of tumor cells to bypass the EGFR pathway and use alternate pathways for cell cycle progression. It is also possible that the presence of *EGFR* mutations may alter the response to EGFR inhibitors, negating the importance of EGFR expression.[21,22] ERBB2 heterodimers are more potent than homodimers and lead to increased and prolonged downstream signaling.[16] Thus, coexpression of ERBB1 and ERBB2 may be more important than the expression of EGFR alone.

The *EGFR* gene is located on chromosome 7 and contains 26 exons.[23] Exons 1 to 14 code for the extracellular domain, exon 15 codes for the transmembrane domain, and exons 16 to 26 code for the intracellular domain. Two enhancer elements, one upstream near the transcription initiation site and the other downstream in intron 1, regulate transcription.[8] The first intron contains a characteristic SSR of repeating (CA) dinucleotides. A polymorphism of 14 to 21 CA dinucleotides is present at a heterozygosity of 72% of Caucasians.[24] A longer sequence of 20 CA repeats is associated with Asian ethnic heritage and a shorter sequence of 16 repeats is associated with Caucasian and African American heritage.[9]

Allele length in breast cancers is inversely proportional to the level of EGFR transcription.[11] Transcription decreased

by approximately 80% when there were 21 repeats, resulting in decreased cellular EGFR protein levels.[10] In NSCLC, squamous cell histology is associated with higher EGFR content,[7] and thus this subtype would be expected to have shorter allele lengths. This was supported by the findings in our study that demonstrated that shorter allele lengths are strongly associated with squamous histology.

The predominant SSR genotype in normal controls, colon cancers, and breast cancers is heterozygous allele lengths of 16/20 repeats.[11,13,24,25] Although the most common genotype in our NSCLC samples was similarly heterozygous, the most common allele lengths were 17/19. The presence of two alleles containing ≥19 CA repeats (which occurred in 10% of tumors) conferred an increased risk of breast cancer development in women with a family history of breast cancer when compared with shorter alleles.[13] In our study, the frequency of two long alleles (≥19 CA) was 39.5%. Because the genotype that confers an increased risk of malignancy in breast cancer occurs even more commonly in the tumors of patients with NCSLC, it raises the question of whether increased lengths of this repeat sequence increase the risk of lung cancer development.

The Hardy-Weinberg formula measures gene frequencies in a reference population. Comparing gene frequencies in a disease-carrying population, such as NSCLC, as was done in the current study, to the reference population is one way of assessing the importance of that gene on the particular disease. The HWE will be maintained if random mating occurs in a large population, there is no migration in or out of the population, and neither mutation nor natural selection occurs. In our study, we found that the alleles were not in HWE. Because ECOG 3590 was a national multicenter trial, it is reasonable to assume that the population involved was large and diverse and that mating was random. Gene mutation is also an unlikely explanation for the lack of HWE, as the SSR polymorphism has been described in both normal controls and in the malignant population. Therefore, natural selection remains the most likely cause of the deviation from HWE. An increase in the frequency of homozygous pairs was seen in our study, suggesting that EGFR homozygous long pairs and short pairs, but not the heterozygous pairs, are important in the development of NSCLC and may confer a survival advantage on the developing neoplasm. In the subset of African Americans patients, the genes were found to be in HWE, suggesting that natural selection of the SSR polymorphism did not occur. However, because the sample size of African Americans in our study was small, these results should be interpreted with caution. Analysis of the *MDR* (multiple drug resistance) gene polymorphism in tumor samples of the same population demonstrated alleles to be in HWE (J. Kolesar et al., unpublished data, 2004). This suggests that loss of HWE did not occur for all genes evaluated in NSCLC tumors but was specific for the *EGFR* gene, supporting the role of EGFR in the development of cancer.

The sum of the CA repeats of both alleles greater than 35 was associated with improved survival. This difference was maintained when patients were randomized by treatment group. A similar cutoff was seen in the preclinical head and neck cancer study, where cell lines with a sum of CA repeats greater than 35 were associated with resistance to erlotinib, suggesting that the polymorphism may be a predictor of response to EGFR inhibitors.[14] In this preclinical study, as seen with other studies, an increase in the length of the polymorphic sequence was associated with decreasing EGFR protein levels. Gene amplification was not seen in these cell lines, eliminating the role of gene amplification in EGFR overexpression, suggesting that EGFR intron 1 polymorphism may play an important role in influencing response to EGFR inhibitors indirectly by regulating transcription of EGFR. Thus, we cannot exclude the possibility that the prognostic ability of the intron 1 polymorphism seen in our study may be related to EGFR levels. Because of tumor sample limitations, EGFR protein levels were not evaluated in our study and should be addressed in a future study. A multivariate analysis in such a study may ascertain the prognostic value of the intron 1 polymorphism, independent of EGFR expression. Because polymorphisms are not unique to tumor tissue, nonmalignant tissue may have sufficed to study this polymorphism as an independent marker. In addition, nonmalignant tissue may have provided a reference for comparison with tumor tissue to assess for potential changes during the malignant process, but the absence of banked matched blood samples did not permit this step. However, our study enabled a unique opportunity to evaluate a biomarker such as the intron 1 polymorphism in tumor tissue in a large cooperative group study with available survival data.

CONCLUSIONS

Distribution of the *EGFR* gene intron 1 polymorphism in stages II and IIIa NSCLC differs from that reported in noncancerous populations and other malignancies. The polymorphism correlates with histology and with prognosis. Further study is needed to detect other potential effects of this sequence. The correlation of intron 1 polymorphism with response to treatment with EGFR inhibitors should be evaluated in future clinical trials. Further study to evaluate loss of heterozygosity in tumor in comparison with reference nonmalignant tissue should be undertaken. The impact of interactions between intron 1 polymorphism and risk factors such as family history and smoking on lung cancer development warrant an exploration.

REFERENCES

1. Jemal A, Siegel R, Murray T, et al. Cancer statistics, 2006. *CA Cancer J Clin.* 2006;56:106–130.
2. Schiller J, Adak S, Feins R, et al. Lack of prognostic significance of p53 and K-ras mutations in primary resected non-small cell lung cancer: results from an ECOG prospective randomized trial of postoperative adjuvant therapy (E3590, INT0015). *J Clin Oncol* 2001;19:448–457.
3. Eberhard DA, Johnson BE, Amler LC, et al. Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. *J Clin Oncol* 2006;23:5900–5909.
4. Khuri FR, Lee JS, Lippman SM, et al. Modulation of proliferating cell nuclear antigen in the bronchial epithelium of smokers. *Cancer Epidemiol Biomarkers Prev* 2001;10:311–318.
5. Franklin WA, Veve R, Hirsch FR, et al. Epidermal growth factor

receptor family in lung cancer and premalignancy. *Semin Oncol* 2002;29 (1 Suppl 4):3–14 .
6. Selvaggi G, Scagliotti GV, Novello S, et al. Prospective evaluation of epidermal growth factor (EGFR) expression in completely resected non-small cell lung cancer (NSCLC): effects of EGFR on long-term follow up (Abstract 1345). *Proc Am Soc Clin Oncol* 2002;21:337a.
7. Hirsch FR, Varella-Garcia M, Bunn PA, et al. Epidermal growth factor receptor in non-small cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* 2003;21:3798–3807.
8. McInerney JM, Wilson MA, Strand KJ, et al. A strong intronic enhancer element of the EGFR gene is preferentially active in high EGFR expressing breast cancer cells. *J Cell Biochem* 2001;80:538–549.
9. Liu W, Innocenti F, Chen P, et al. Interethnic difference in the allelic distribution of human epidermal growth factor receptor intron 1 polymorphism. *Clin Cancer Res* 2003;9:1009–1012.
10. Gebhardt F, Zanker KS, Brandt B. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. *J Biol Chem* 1999;274:13176–13180.
11. Buerger H, Gebhardt F, Schmidt H, et al. Length and loss of heterozygosity of an intron 1 polymorphic sequence of egfr is related to cytogenetic alterations and epithelial growth factor. *Cancer Res* 2000; 60:854–857.
12. Keller SM, Adak S, Wagner H, et al. A randomized trial of postoperative adjuvant therapy in patients with completely resected stage II or IIIa non-small cell lung cancer. *N Engl J Med* 2000;343:1217–1222.
13. Brandt B, Hermann S, Straif K, et al. Modification of breast cancer risk in young women by a polymorphic sequence in the egfr gene. *Cancer Res* 2004;64:7–12.
14. Amador ML, Oppenheimer D, Perea S, et al. An epidermal growth factor receptor intron 1 polymorphism mediates response to epidermal growth factor receptor inhibitors. *Cancer Res* 2004;64:9139–9143.
15. Breiman L, Friedmand JH, Olshen RA, et al. Classification and Regression Trees. Belmont, CA: Wadsworth International Group, 1984.
16. Yarden Y, Sliwkowski MX. Untangling the ErbB signaling network. *Nat Rev Mol Cell Biol* 2001;2:127–137.
17. Kari C, Chan T, Rocha de Quadros M, Rodeck U. Targeting the epidermal growth factor receptor in cancer. *Cancer Res* 2003;63:1–5.
18. Andrulis IL, Bull SB, Blackstein ME, et al. Neu/erbB-2 amplification identifies a poor-prognosis group of women with node-negative breast cancer. Toronto Breast Cancer Study Group. *J Clin Oncol* 1998;16: 1340–1349.
19. Vogel CL, Cobleigh MA, Tripathy D, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20:719–726.
20. Fukuoka M, Kris M, Giaccone G, et al. Phase II trials of gefitnib ("Iressa", ZD1839): rapid and durable objective responses in patients with advanced non-small cell lung cancer (IDEAL 1 and IDEAL 2) (Abstract P615). *Lung Cancer* 2003;41 (Suppl 2):S247.
21. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139.
22. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304: 1497–1500.
23. Haley J, Whittle N, Bennet P, et al. The human EGF receptor gene: structure of the 110 kb locus and identification of sequences regulating its transcription. *Oncogene Res* 1987;1:375–396.
24. Chi DD, Hing AV, Helms C, et al. Two chromosome 7 dinucleotide repeat polymorphisms at gene loci epidermal growth factor receptor (EGFR) and pro alpha 2 (I) collagen (COL1A2). *Hum Mol Genet* 1992;1:135.
25. McKay JA, Murray LJ, Curran S, et al. Evaluation of the epidermal growth factor receptor (EGFR) in colorectal tumours and lymph node metastases. *Eur J Cancer* 2002;38:2258–2264.